TECHNIQUES AND METHODS OF GIS FOR REGIONAL DEVELOPMENT

Abstract
The paper presents some techniques and methods of geographic informatic systems, adapted and applied to the regional development applications. The characteristics of spatial databases, as they are required for regional applications are presented in the first part of the paper. Methods of displaying spatial data are compared and GIS procedures are described. The main techniques available in the GIS and useful for regional data visualization are exemplified. It is presented a case study that illustrates the use of GIS for analyzing the concentration of economic activities, based on employment datasets, for various branches in Romania's counties.

Keywords: regional development, GIS, regional data visualisation, cartogram, spatial analysis

JEL CODES: R0, R12

TEHNICI ŞI METODE ALE SISTEMELOR INFORMATICE GEOGRAFICE PENTRU DEZVOLTAREA REGIONALĂ

Adriana REVEIU
Associated Professor Ph.D., Economic Cybernetics, Statistics and Informatics Faculty, Bucharest Academy of Economic Studies
E-mail: reveiua@ase.ro

Rezumat
Lucrarea prezintă câteva tehnici şi metode specifice sistemelor informatice geografice, adaptate şi utilizate în aplicaţii din domeniul dezvoltării regionale. În prima parte a lucrării sunt prezentate caracteristicile bazelor de date spaţiale, necesare problematicii specifice aplicaţiilor regionale. Sunt prezentate comparativ metodele şi procedurile GIS de afişare a datelor regionale. Sunt descrise şi exemplificate principalele tehnici disponibile în GIS pentru vizualizarea datelor spaţiale. In ultima parte a lucrării este prezentat un studiu de caz care ilustrează folosirea metodelor şi tehnicilor GIS pentru analiza şi vizualizarea concentraţiei şi distribuţiei activităţilor în diversele ramuri economice, în judeţele României.

Cuvinte cheie: Dezvoltare regională, sisteme informatice geografice, vizualizarea datelor regionale, cartogramă, analiză spaţială

1. INTRODUCTION

In recent years, Geographic Information Systems (GIS) have become an important tool for regional research, as about 80-90% of data collected and used for regional and environmental information systems are related to geography. (Huxhold, 1991)

The huge amount of spatial data generated by GIS expansion, the increasing number of geographic informatic applications available, the computerization of a large amount of information sources, and the availability of digital map has increased the opportunity and need for the usage of methods and techniques for spatial data classification and integration with socio-economical data, for both research and applied purposes.

The main problems identified when we try to classify spatial data are: there is a large number of types of administrative, cultural, social areas; there is a large number of types of variables available for geographic and socio-economical data representation; specifically data has non-normal variable distributions, because most geographic data usually have very complex frequency distributions; usually there are non linear relationships between various variable in analysis; between analysed variable there is spatial dependency; the uncertainty of economic, social and geographic data; systematic non random variations in spatial representation.

GIS provides: an integrated computing environment for economic, social and geographical data; a set of analytical capabilities to operate on topological relationships and on the spatial aspects of the geographical data; a set of interactive and programming tools to manage the non-spatial attributes of economic and social data; and tools to combine the non-spatial and spatial attributes of geographical data. GIS facilitates the integration of distinct data sets, creation of new data sets, development and analysis of spatially variables. GIS is associated with the process of capturing information from spatial data. There are two important issues related with the process of data capture: the information acquisition and the information accuracy. There are two main techniques to turn data into information: first of all the visualisation of economic and social information on the maps and the statistical and mathematical modelling of spatial and non-spatial attributes of data.

There is a vast set of techniques that have been developed to visualize spatial data and it remains a very intense and rich area of research. GIS based visualization of territorial data allows the user to virtually interact with the real world with and immersion into artificial worlds.

PROCEEDINGS OF THE SEVENTH ADMINISTRATION AND PUBLIC MANAGEMENT INTERNATIONAL CONFERENCE
"State Reform: Public Administration and Regional Development"
Bucharest, June 21-22, 2011

The basic elements of a successful visualization and usage of geographical data is the understanding of how human perceive GIS models usage, how people think about space and time, and how spatial environments might be better represented using computers and digital data.

2. TECHNIQUES FOR REGIONAL DATA VISUALIZATION IN GIS

Efficient visualization of geographical data can be a challenge. Usually, colour and shadows are used for information sharing and to figure the differences between regions. The thematic maps on various topics are commonly used and extremely widespread throughout the world, like social, political, economics and so on. The most common used cartogram, on a social issue, is designed to represent the population density, emphasizing the proper volume and population distribution in the territory and not the size of the region. Representation of election results using cartograms is an example of political cartogram, each region being sized according to the share of the votes in total votes. On trade issues, the cartograms are used to represent production and consumption of certain products in various markets.

Data visualization can greatly enhance the understanding of multivariate data structures. The socio-economical regional data distribution can be visualized by projecting the data into two-dimensional space. The membership of one region to one category is usually represented by different colours or by dividing regions into several panels of a grille display (Becker, Cleveland and Shyu, 1996). In addition, outline plots provide a popular tool for diagnosing the quality of a partition. Some of the popularity of self-organizing feature maps with practitioners in various fields can be explained by the fact that the results can be "easily" visualized (Kohonen, 1989).

The common of modern GIS applications are characterized by sophisticated graphics and this allows GIS to provide effective support for problem of geographical data representation in space.

Cartography is an important source of principles for the design of business graphics (Speier and Morris, 2003). Information visualization techniques have been widely applied in science and geography, but only recently have been integrated into business applications (Speier and Morris, 2003).

The complex nature of spatial data requires GIS to use sophisticated visualization techniques to represent information. It is therefore quite challenging for GIS to provide an interactive interface on the same screen.

Visualization has been recognized in the GIS community as an important aspect of GIS (Buttenfield and Mackaness, 1991). This may reflect support for the map view of GIS. One limitation of GIS interface designs is that they are seen to provide a means for visualizing results only, rather than providing a comprehensive problem representation for all stages of the problem (Blaser, Sester and Egenhofer, 2000).

There are a lot of techniques and tools used for data visualisation in GIS. In this paper I present the two most important tools required for regional application development using GIS support. The most important tools and techniques available for regional data representation are cartograms and spatial analysis.

Cartograms are unique representations of geographical space. A cartogram is a type of graphic that describes attributes of geographic objects as the object's area. Mapping requirements include the preservation of shape, orientation contiguity, and the data that have suitable variation. Successful communication depends on how well the map reader recognizes the shapes of the internal enumeration units, the accuracy of estimating these areas, and effective legend design. Because a cartogram does not describe geographic space, but rather changes the size of objects depending on a certain attribute, a cartogram is not a true map. Cartograms vary on their degree in which geographic space is changed: some appear very similar to a map; however some don't look like a map (Borden, 1996).

There are three main types of cartograms, each having a very different way of showing attributes of geographic objects: non-contiguous, contiguous and Dorling cartograms.

A **non-contiguous cartogram** is the simplest and easiest type of cartogram. In a non-contiguous cartogram, the geographic objects do not have to maintain connectivity with their adjacent objects. This connectivity is called topology. By freeing the objects from their adjacent objects, they can grow or get smaller in size and still maintain their shape (Borden, 1996).

In a **contiguous cartogram** is maintained the true topology, the objects remain connected with each other, but this causes deformation in shape (Mackaness, 1996). In this case, the cartographer must make the objects of the appropriate size, to represent the attribute value, but she or he must also maintain the shape of objects as best as possible, so that the cartogram can be easily interpreted.

A **Dorling cartogram** maintains neither shape topology nor object centroids, though it has proven to be a very effective cartogram method. This type of cartogram was named after its inventor, Danny Dorling of the University of Leeds. To create a Dorling cartogram, instead of expanding or shrinking the objects themselves, the cartographer will replace the objects with a uniform shape, usually a circle, of the appropriate size (Dorling, 1995).

PROCEEDINGS OF THE SEVENTH ADMINISTRATION AND PUBLIC MANAGEMENT INTERNATIONAL CONFERENCE
"State Reform: Public Administration and Regional Development"
Bucharest, June 21-22, 2011

Spatial analysis manipulates geographic coordinates and associated attribute data for the purposes of solving a spatial problem. In this paper, spatial analyses that are especially relevant for regional development applications are illustrated. Many existing attributes of economic databases are spatial, including addresses of companies' headquarters, telephone area codes, and statistical data related to the regions or counties. These are not typically thought of as spatial attributes because they are not specified as latitude and longitude coordinates. Typically, what makes a database spatial is the connection of the data to a geographically referenced coordinate system. A geographical coordinate system exactly locates objects on the earth in terms of an X and Y coordinate position. Latitude and longitude are the most frequently used reference coordinates. Both are measured as angles from the centre of the earth as a point to a point on the surface of the earth. Many GIS databases are geographically referenced with transformed coordinates from a different map projection and associated coordinate system and are typically referred to as X and Y coordinates (Blaser, Sester and Egenhofer, 2000). Commercially available DataBase Management Systems (DBMS), like Oracle, Informix and ESRI's Spatial Data Engine have, directly or through extensions, implemented support for spatial data.

The **shape files** have become an industry standard for storing some types of spatial data. Spatial data requires a unique set of operations to manipulate the X and Y coordinates stored in the DBMS. One technique that can be used to transform the location of one building into spatial data represented as an X and Y location is to perform geocoding. Geocoding or address matching is a process that compares two addresses to determine whether they are the same (DeSanctis,1984). To match addresses, the GIS software examines the components of addresses in the database file where addresses are maintained and in the attribute table associated with a GIS layer of roads.

3. SPATIAL DISPLAY OF DATA

The benefit of a database that stores its economic statistical data in a GIS is that it allows visualizing the spatial model of those regions or counties. In the case of a large database with economic statistical data for Romania's counties, collected over the years, the value of the database is not its large size, but its capacity to answer questions, to structure the queries, and the methods available for displaying the results of the queries on the map. Alternatively, it is possible to display the results using various types of charts. It would also be useful to display data on a map, which is made possible when the data are stored in a GIS's database. Such a map might include, for example, the distribution of employment or of companies' income around the Romania's counties or regions.

An important component included in GIS maps is the legend, which containes the appropriate symbols, colours, and classifications used for drawing the map layers. For example, a map illustrating the average gross nominal monthly earnings at territorial level patterns may vary the symbol size for levels of earnings or it is possible to symbolise various levels of earnings using different colours or a colour ramp. A **colour ramp** is a spectrum of colour that allows a distinct colour for each represented value. As an example, in figure 1, I picked a red to yellow colour ramp that shows a spectrum of colours between yellow to red, yellow being associated with the lower values of the calculated index and red with highest values. The GIS represents the values of location quotient, which is discussed in the next section.

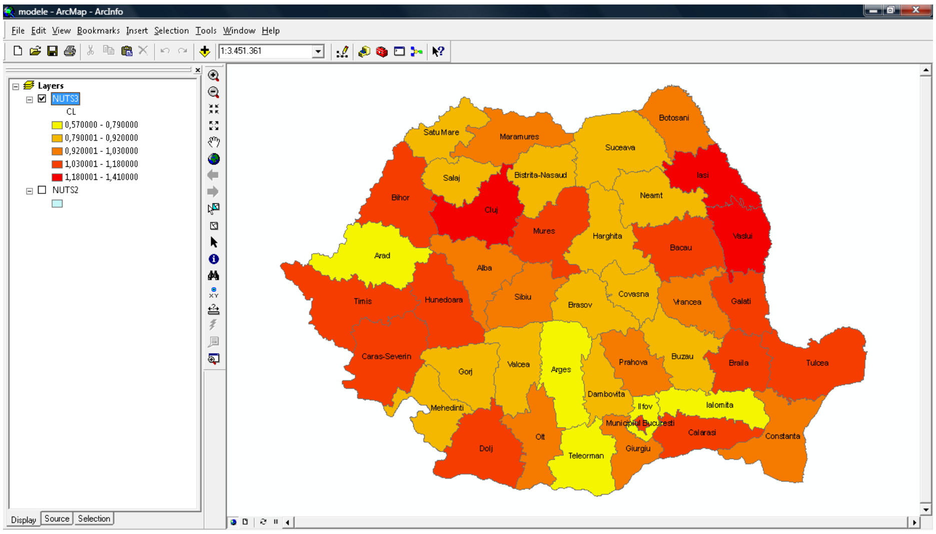

FIGURE 1 - USING A COLOR RAMP TO REPRESENT THE DISTRIBUTION OF ACTIVITY RELATED TO HEALTH AND SOCIAL ASSISTANCE BRANCH IN ROMANIA

Another symbolization technique is a **dot density map**. This technique is performed by associate the value of calculated index, as in the example from figure 2, to a number of dots that are shown within each region area, each dot being correlate with a specific value displayed in the legend, chooses function on the range of represented values.

PROCEEDINGS OF THE SEVENTH ADMINISTRATION AND PUBLIC MANAGEMENT INTERNATIONAL CONFERENCE
"State Reform: Public Administration and Regional Development"
Bucharest, June 21-22, 2011

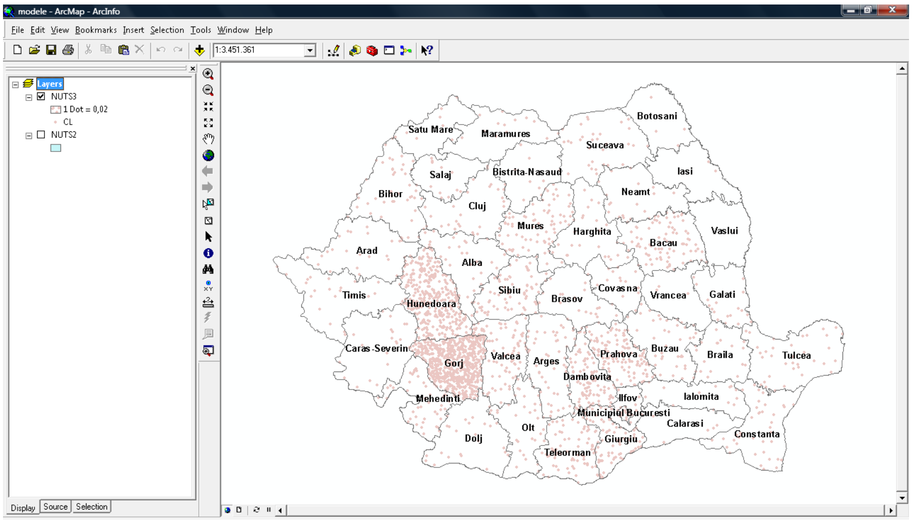

FIGURE 2 – USING A DOT DENSITY MAP TO REPRESENT THE DISTRIBUTION OF ACTIVITY RELATED TO *MINING AND QUARRYING* BRANCH IN ROMANIA

GIS software uses a lot of classification methods to display the values on the map. When we use the symbolization techniques, we have to decide the number of classes needed and how to break the data into classes. Each method uses a math equation to calculate the range of each class. In this paper there are presented some of the more common methods employed in GIS software.

- **Natural breaks** are used when there are large jumps between values of observations to be represented. The natural breaks method looks for evident breaks or gaps in the data, to establish classes. If five classes are requested, for example, the GIS will try to find five areas that are separated by a gap in the clusters of values.
- In the **equal interval classification** the lowest value is subtracted from the highest value to compute a range, and then the range is divided by the number of classes that are desired. The resulting number is then added to the lowest value to get the upper range for the lowest class. The resulting number is then added to that, to get the upper range for the second class, and so on. In the examples presented in figure 1 and 3, the equal interval classification is used to split the values level of the represented index.
- An **equal area classification** sets class boundaries so as to include an equal proportion of a map area into each established class. The map will appear balanced in that each class will represent approximately the same area in extent.
- **Standard deviation**, the average difference of the set of values from the mean of the set of values, is used by the GIS to establish a classification. The formula to compute a standard

deviation is: $s = \sqrt{\frac{\sum_{i=1}^{n}\left(x_i - \bar{x}\right)^2}{n}}$, where s is the standard deviation, n is the number of items in the list, x is the value of an item, $\bar{x}$ is the mean of the items. For classification breaks, the user selects the number of standard deviations, for instance two will generate four classes, two above the mean and two below the mean. This is an effective method for showing extremes in the data: as in the case of average gross nominal monthly earnings at counties level, it is possible to quickly visualize extreme low and high values.

4. GIS ANALYSIS OF INDUSTRIAL SPECIALIZATION IN ROMANIA

Regional development specific applications need to know the availability of types of employment appropriate to their workplaces. This information is useful as one factor in deciding on the sitting of facilities, and for existing facilities in determining the costs of hiring. GIS spatial analysis can identify the intensity distributions of employment for industrial specializations. The location of business activities may appear widely dispersed.

To exemplify the distribution and concentration of economic activity and to display on the map the calculated data, I implemented the location quotient method. A location quotient expresses the share of employment in a given industry from a county or region as a percentage of the share of employment in the same industry within all the country.

Location quotient method is designed to group local industries into clusters, using regional data about employees. In order to identify the leading regions of a spatially concentrated industry Kim (1995) and Hoover (1936), suggest to calculate for each locational unit in a given sample industries' employment shares, with respect to each industry's total employment in the aggregated locational unit.

$$Location\ quotient = \frac{\frac{n_{A,R}}{N_R}}{\frac{n_{A,T}}{N_T}}$$

, where: $n_{A,R}$- the number of employees in industry A, in region R, N_R - the whole number of employees, in the region R, $n_{A,T}$- the number of employees, in industry A, at the national level, N_T- the whole number of employees, from national level.

A region is considered to be specialized in one industry if the location quotient calculated for that region is greater than or equal to 1.5. The method has been applied in many countries because it uses employment data only, which are relatively easily available. The main shortcoming of the method is the large dependence by the regions bounds choosing. The choice of regions must be a priori to identify clusters.

In the figure 3, I presented the resulted map of location quotient method for Agriculture, forestry and fishing branch. The specialised counties of Romania in Agriculture, forestry and fishing branch are coloured with lavender and non-specialised counties with yellow. Using the results provided by the location quotient method, we can identify two possible clusters in this branch, the first in the West part of Romania (Botosani, Suceava, Neamt, Vaslui, and Vrancea counties) and the second in the South part (Mehedinti, Olt, Teleorman, Giurgiu, Calarasi, and Ialomita counties).

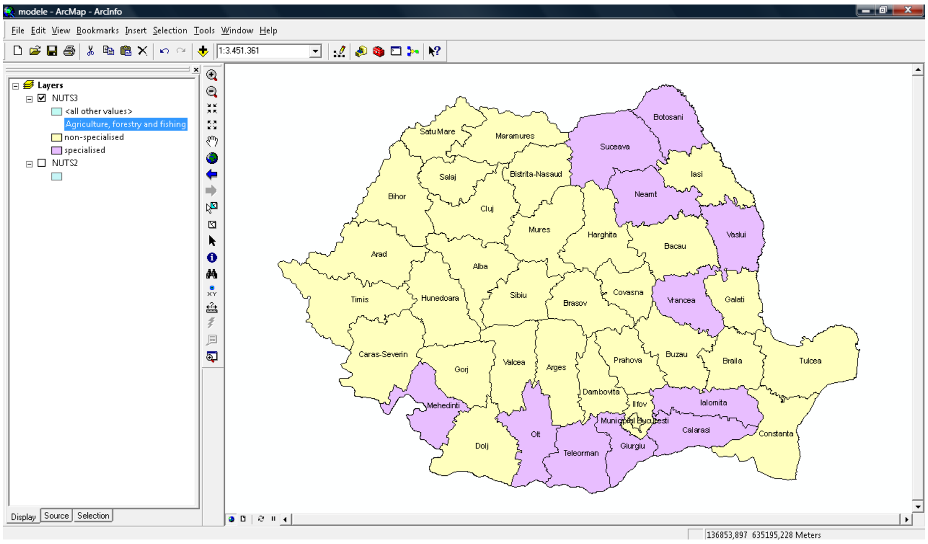

FIGURE 3 – GIS REPRESENTATION OF THE ECONOMIC DISTRIBUTION OF *AGRICULTURE, FORESTRY AND FISHING* BRANCH IN ROMANIA

5. CONCLUSIONS

There are numerous GIS techniques and methods available for regional applications. Many of techniques and methods presented are used in many different types of GIS applications. In this paper I exemplyfied there usage on a case study specific for regional development area. The application of GIS for analyzing the concentration and distribution of economic branches in Romania shows the benefits of linking statistical economic data to a GIS map used to display the results in the territory.

The spatial relationships revealed through graduated symbol mapping of locations quotients can yield better understanding of the diversity of the employment concentration in Romania's counties.

ACKNOWLEDGMENTS

This work was supported by the project "Performance and excellence in postdoctoral research in Romanian economics science domain", contract POSDRU/89/1.5/S/59184.

REFERENCES

Becker, R., Cleveland, W. and Shyu, M.-J. (1996). The visual design and control of trellis display, *Journal of Computational and Graphical Statistics* 5, pp. 123-155.

Blaser, A.D., Sester, M. and Egenhofer, M.J. (2000). Visualization in an early stage of the problem-solving process in GIS, *Computers & Geosciences*, 26(1), pp. 57-66.

Borden, D.D. (1996). *Cartography: Thematic Map Design*, 4th edition, William C Brown Publishing House.

Buttenfield, B.P. and Mackaness, W.A. (1991). Geographical Information Systems, Volume 1: Principles, Harlow, Essex, *UK: Longman Scientific & Technica*, pp. 427-443.

DeSanctis, G. (1984). Computer graphics as decision aids: Directions for research. *Decision Sciences*, 15(4), pp. 463-487.

Dorling, D. (1995). *Area Cartograms: Their Use and Creation*, Department of Geography, University of Newcastle upon Tyne, England.

Hoover (1936). The Measurement of Industrial Localization, *The Review of Economic and Statistics*, 18(4), pp. 162—171.

Huxhold, W. (1991). *An introduction to urban geographic information systems*, Oxford University Press.

Kim, S. (1995). Expansion of Markets and the Geographic Distribution of Economic Activities: The Trends in U. S. Regional Manufacturing Structure, 1860-1987. *The Quarterly Journal of Economics,* 110(4), pp. 881-908.

Kohonen, T. (1989). *Self-organization and Associative Memory*, 3rd edn, Springer, New York.

Mackaness, W. (1996). *Automated Cartography and the Human Paradigm*, Cartographic Design, Theoretical and Practical Perspectives. Chichester NY: John Wiley & Sons Ltd.

Speier, C. and Morris, M.G. (2003). The influence of query interface design on decisionmaking Performance, *MIS Quarterly*, 27(3), pp. 397-423.